Canonical correlation analysis (I)

Reading:
AMSA: pages 539-549

Multivariate Analysis, Spring 2016
Institute of Statistics, National Chiao Tung University

May 24, 2016

Brief outline

1. Introduction
2. Canonical variates and canonical correlations
3. Interpreting the population canonical variables

Introduction

- Canonical correlation analysis seeks to identify and quantify the associations between **two sets** of variables.
  - For example, relating government policy variables with economic goal variables,
  - Relating college "performance" variables with precollege "achievement" variables.

- Canonical correlation analysis focuses on the correlation between a **linear combination** of the variables in one set and a **linear combination** of the variables in another set.
  - First determine the pair of linear combinations having the largest correlation.
  - Next determine the pair of linear combinations having the largest correlation among all pairs uncorrelated with the initially selected pair, and so on.
- The pairs of linear combinations are called the **canonical variables**, and their correlations are called **canonical correlations**.

Variables considered

- Interest in measures of association between two groups of variables: $\mathbf{X}^{(1)}_{(p\times 1)}$ and $\mathbf{X}^{(2)}_{(q\times 1)}$ with $p \leq q$.
- Let

$$\mathrm{E}(\mathbf{X}^{(1)}) = \boldsymbol{\mu}^{(1)}; \qquad \mathrm{Cov}(\mathbf{X}^{(1)}) = \boldsymbol{\Sigma}_{11}$$

$$\mathrm{E}(\mathbf{X}^{(2)}) = \boldsymbol{\mu}^{(2)}; \qquad \mathrm{Cov}(\mathbf{X}^{(2)}) = \boldsymbol{\Sigma}_{22}$$

$$\mathrm{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^{T}$$

- Consider $\mathbf{X}^{(1)}_{(p\times 1)}$ and $\mathbf{X}^{(2)}_{(q\times 1)}$ jointly

$$\mathbf{X}_{((p+q)\times 1)} = \left[\begin{array}{c} \mathbf{X}^{(1)} \\ \hline \mathbf{X}^{(2)} \end{array}\right] = \left[\begin{array}{c} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_p^{(1)} \\ \hline X_1^{(2)} \\ X_2^{(2)} \\ \vdots \\ X_q^{(2)} \end{array}\right]$$

with mean vector

$$\boldsymbol{\mu}_{((p+q)\times 1)} = \mathrm{E}(\mathbf{X}) = \left[\begin{array}{c} \mathrm{E}(\mathbf{X}^{(1)}) \\ \hline \mathrm{E}(\mathbf{X}^{(2)}) \end{array}\right] = \left[\begin{array}{c} \boldsymbol{\mu}^{(1)} \\ \hline \boldsymbol{\mu}^{(2)} \end{array}\right]$$

and covariance matrix

$$\boldsymbol{\Sigma}_{((p+q)\times(p+q))} = \mathrm{E}(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T$$

$$= \left[\begin{array}{c|c} \mathrm{E}(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})^T & \mathrm{E}(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})^T \\ \hline \mathrm{E}(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})^T & \mathrm{E}(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})^T \end{array}\right]$$

$$= \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{11\ (p\times p)} & \boldsymbol{\Sigma}_{12\ (p\times q)} \\ \hline \boldsymbol{\Sigma}_{21\ (q\times p)} & \boldsymbol{\Sigma}_{22\ (q\times q)} \end{array}\right]$$

- The covariances between pairs of variables from different sets $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are contained in $\boldsymbol{\Sigma}_{12}$.
- When p and q are relatively large, interpreting the elements of $\boldsymbol{\Sigma}_{12}$ collectively is ordinarily hopeless.
- The main task of canonical correlation analysis is to summarize the association between the $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ sets in terms of a **few** carefully chosen covariances (or correlations) between the **linear combinations** of the variables in $\mathbf{X}^{(1)}$ and the **linear combinations** of of the variables in $\mathbf{X}^{(2)}$.

Canonical variates and canonical correlations

- Set linear combinations

$$U = \mathbf{a}^T \mathbf{X}^{(1)} \qquad V = \mathbf{b}^T \mathbf{X}^{(2)}$$

- We can obtain

$$\begin{aligned}
\mathrm{Var}(U) &= \mathbf{a}^T \mathrm{Cov}(\mathbf{X}^{(1)})\mathbf{a} = \mathbf{a}^T \mathbf{\Sigma}_{11}\mathbf{a} \\
\mathrm{Var}(V) &= \mathbf{b}^T \mathrm{Cov}(\mathbf{X}^{(2)})\mathbf{b} = \mathbf{b}^T \mathbf{\Sigma}_{22}\mathbf{b} \\
\mathrm{Cov}(U, V) &= \mathbf{a}^T \mathrm{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})\mathbf{b} = \mathbf{a}^T \mathbf{\Sigma}_{12}\mathbf{b}
\end{aligned}$$

- We seek coefficient vectors **a** and **b** such that

$$\mathrm{Corr}(U, V) = \frac{\mathbf{a}^T \mathbf{\Sigma}_{12}\mathbf{b}}{\sqrt{\mathbf{a}^T \mathbf{\Sigma}_{11}\mathbf{a}}\sqrt{\mathbf{b}^T \mathbf{\Sigma}_{22}\mathbf{b}}} \tag{1}$$

is as large as possible.

- Define

$$
\begin{aligned}
\text{1st canonical variate pair} &= \text{the pair of linear combinations } U_1 \text{ and } V_1 \text{ having unit variances, which maximize the correlation (1)} \\
\text{2nd canonical variate pair} &= \text{the pair of linear combinations } U_2 \text{ and } V_2 \text{ having unit variances, which maximize the correlation (1) among all choices that are uncorrelated with the 1st canonical variate pair} \\
\vdots \quad & \vdots \quad \vdots \\
k\text{th canonical variate pair} &= \text{the pair of linear combinations } U_k \text{ and } V_k \text{ having unit variances, which maximize the correlation (1) among all choices uncorrelated with the previous } k-1 \text{ canonical variate pairs}
\end{aligned}
$$

- Let $\rho_1^{*2} \geq \rho_2^{*2} \geq \cdots \geq \rho_p^{*2}$ be the eigenvalues of $\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2}$, and $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_p$ are the associated $(p \times 1)$ eigenvectors.

  $\rho_1^{*2}, \rho_2^{*2}, \cdots, \rho_p^{*2}$ are also the p largest eigenvalues of $\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1/2}$ with corresponding $(q \times 1)$ eigenvectors $\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_p$.

  Note that $\mathbf{f}_k = (1/\rho_i^*)\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{e}_k, k = 1, \cdots, p.$

- Result:

$$\max_{\mathbf{a},\mathbf{b}} \mathrm{Corr}(U, V) = \rho_1^*$$

attained by the linear combinations (1st canonical variate pair)

$$U_1 = \underbrace{\mathbf{e}_1^T \mathbf{\Sigma}_{11}^{-1/2}}_{\mathbf{a}_1^T} \mathbf{X}^{(1)} \text{ and } V_1 = \underbrace{\mathbf{f}_1^T \mathbf{\Sigma}_{22}^{-1/2}}_{\mathbf{b}_1^T} \mathbf{X}^{(2)}$$

The kth canonical variate, $k = 2, 3, \cdots, p$,

$$U_k = \mathbf{e}_k^T \mathbf{\Sigma}_{11}^{-1/2} \mathbf{X}^{(1)} \text{ and } V_k = \mathbf{f}_k^T \mathbf{\Sigma}_{22}^{-1/2} \mathbf{X}^{(2)}$$

maximizes

$$\mathrm{Corr}(U_k, V_k) = \rho_k^*$$

among those linear combinations uncorrelated with the preceding $1, 2, \cdots, k-1$ canonical variate pairs.

- The canonical variates have the properties

$$\begin{aligned}
\mathrm{Var}(U_k) &= \mathrm{Var}(V_k) = 1 \\
\mathrm{Cov}(U_k, U_\ell) &= \mathrm{Corr}(U_k, U_\ell) = 0 \quad k \neq \ell \\
\mathrm{Cov}(V_k, V_\ell) &= \mathrm{Corr}(V_k, V_\ell) = 0 \quad k \neq \ell \\
\mathrm{Cov}(U_k, V_\ell) &= \mathrm{Corr}(U_k, V_\ell) = 0 \quad k \neq \ell
\end{aligned}$$

for $k, \ell = 1, 2, \cdots, p$.

Canonical variates from standardized variables

- The original variables are standardized with $\mathbf{Z}^{(1)} = [Z_1^{(1)}, Z_2^{(1)}, \cdots, Z_p^{(1)}]^T$ and $\mathbf{Z}^{(2)} = [Z_1^{(2)}, Z_2^{(2)}, \cdots, Z_q^{(2)}]^T$.
- Here, $\mathrm{Cov}(\mathbf{Z}^{(1)}) = \boldsymbol{\rho}_{11}$, $\mathrm{Cov}(\mathbf{Z}^{(2)}) = \boldsymbol{\rho}_{22}$, and $\mathrm{Cov}(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}) = \boldsymbol{\rho}_{12} = \boldsymbol{\rho}_{21}^T$.
- $\mathbf{e}_k$ and $\mathbf{f}_k$ are the eigenvectors of $\boldsymbol{\rho}_{11}^{-1/2}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1/2}$ and $\boldsymbol{\rho}_{22}^{-1/2}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1/2}$, respectively.

  $\rho_1^{*2} \geq \rho_2^{*2} \geq \cdots \geq \rho_p^{*2}$ are the eigenvalues of $\boldsymbol{\rho}_{11}^{-1/2}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1/2}$ (or, equivalently, the largest eigenvalues of $\boldsymbol{\rho}_{22}^{-1/2}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1/2}$).

- The canonical variates are of the form

$$U_k = \mathbf{a}_k^T \mathbf{Z}^{(1)} = \mathbf{e}_k^T \boldsymbol{\rho}_{11}^{-1/2} \mathbf{Z}^{(1)}$$

$$V_k = \mathbf{b}_k^T \mathbf{Z}^{(2)} = \mathbf{f}_k^T \boldsymbol{\rho}_{22}^{-1/2} \mathbf{Z}^{(2)}$$

The canonical correlations, ρ_k^*, satisfy

$$\mathrm{Corr}(U_k, V_k) = \rho_k^*, \quad k = 1, 2, \cdots, p$$

Canonical coefficients: original vs. standardized

- Notice that

$$\mathbf{a}_k^T(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)}) = \mathbf{a}_k^T \mathbf{V}_{11}^{1/2} \mathbf{Z}^{(1)}$$

  where $\mathbf{V}_{11}$ is the diagonal matrix with ith element $\mathrm{Var}(X_i^{(1)})$.
- If $\mathbf{a}_k^T$ is the coefficient vector for the kth canonical variate from $\mathbf{X}^{(1)}$, then $\mathbf{a}_k^T \mathbf{V}_{11}^{1/2}$ is the coefficient vector for the kth canonical variate from the standardized variables $\mathbf{Z}^{(1)}$.
- Similarly, if $\mathbf{b}_k^T$ is the coefficient vector for the kth canonical variate from $\mathbf{X}^{(2)}$, then $\mathbf{b}_k^T \mathbf{V}_{22}^{1/2}$ is the coefficient vector for the kth canonical variate from the standardized variables $\mathbf{Z}^{(2)}$.
- The canonical correlations are **unchanged** by the standardization.

- Due to the special structure of the matrix

$$\mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1/2} \text{ or } \boldsymbol{\rho}_{11}^{-1/2}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1/2},$$

  above relationship is unique to canonical correlation analysis.
  - For example, this is not true for principal component analysis.

Interpreting the population canonical variables

- Canonical variables are, in general, artificial; that is, they have no physical meaning.
- However, their meanings can be **identified** in terms of original variables.
  - Identification can be through the canonical coefficients $\mathbf{a}_k$, $\mathbf{b}_k$: the same as in principal component analysis.
  - Identification is aided by computing the correlations between the canonical variates and the original variables.
  - Above correlations must be interpreted with caution: they provide only univariate information, in the sense that they do not indicate how the original variables contribute **jointly** to the canonical variables.

- Let $\mathbf{A}_{(p\times p)} = [\mathbf{a}_1, \cdots, \mathbf{a}_p]^T$ and $\mathbf{B}_{(q\times q)} = [\mathbf{b}_1, \cdots, \mathbf{b}_q]^T$, so that the vector of canonical variables are

$$\mathbf{U}_{(p\times 1)} = \mathbf{A}\mathbf{X}^{(1)} \quad \mathbf{V}_{(q\times 1)} = \mathbf{B}\mathbf{X}^{(2)},$$

  where we are primarily interested in the first p canonical variables in $\mathbf{V}$.

- Then, since $\mathrm{Var}(U_k) = 1, k = 1, \cdots, p$,

$$\begin{aligned}\boldsymbol{\rho}_{\mathbf{U},\mathbf{X}^{(1)}} &= \mathrm{Corr}(\mathbf{U}, \mathbf{X}^{(1)}) = \mathbf{I}^{-1/2}\mathrm{Cov}(\mathbf{U}, \mathbf{X}^{(1)})\mathbf{V}_{11}^{-1/2} \\ &= \mathbf{A}\boldsymbol{\Sigma}_{11}\mathbf{V}_{11}^{-1/2}\end{aligned}$$

- Similar calculations yield

$$\begin{aligned}\boldsymbol{\rho}_{\mathbf{U},\mathbf{X}^{(1)}} = \mathbf{A}\boldsymbol{\Sigma}_{11}\mathbf{V}_{11}^{-1/2} &\quad \boldsymbol{\rho}_{\mathbf{V},\mathbf{X}^{(2)}} = \mathbf{B}\boldsymbol{\Sigma}_{22}\mathbf{V}_{22}^{-1/2} \\ \boldsymbol{\rho}_{\mathbf{U},\mathbf{X}^{(2)}} = \mathbf{A}\boldsymbol{\Sigma}_{12}\mathbf{V}_{22}^{-1/2} &\quad \boldsymbol{\rho}_{\mathbf{V},\mathbf{X}^{(1)}} = \mathbf{B}\boldsymbol{\Sigma}_{21}\mathbf{V}_{11}^{-1/2} \qquad (2)\end{aligned}$$

- For canonical variables derived from standardized variables, we have

$$\rho_{\mathbf{U},\mathbf{Z}^{(1)}} = \mathbf{A_Z}\rho_{11} \quad \rho_{\mathbf{V},\mathbf{Z}^{(2)}} = \mathbf{B_Z}\rho_{22}$$

$$\rho_{\mathbf{U},\mathbf{Z}^{(2)}} = \mathbf{A_Z}\rho_{12} \quad \rho_{\mathbf{V},\mathbf{Z}^{(1)}} = \mathbf{B_Z}\rho_{21} \tag{3}$$

  where $\mathbf{A_Z}$ and $\mathbf{B_Z}$ are the matrices whose rows contains the canonical coefficients for the $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$ sets, respectively.
- Notice that the correlations in the matrices in (2) have the **same** numerical values as those appearing in (3).
- See Example 10.2 of Johnson & Wichern (2007) (page 546).